

AD-A069 816

UNION COLL AND UNIV SCHENECTADY NY INST OF ADMINISTR--ETC F/G 12/1  
A SIMPLE METHOD FOR REGRESSION ANALYSIS WITH CENSORED DATA.(U)

APR 79 J SCHMEE, 6 J HAHN

N00014-77-C-0438

NL

UNCLASSIFIED

1 OF 1

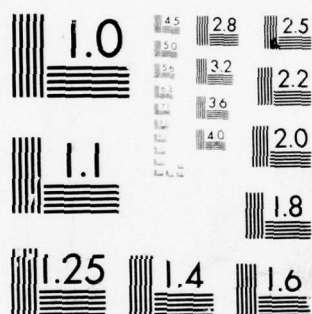
AD  
A069816



END  
DATE  
FILMED

7-79

DDC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 069816

LEVEL

12



A SIMPLE METHOD FOR REGRESSION ANALYSIS  
WITH CENSORED DATA

DDC FILE COPY



This document has been approved  
for public release and sale; its  
distribution is unlimited.

INSTITUTE OF ADMINISTRATION AND MANAGEMENT  
Union College and University

79 06 12 170

12

6

A SIMPLE METHOD FOR REGRESSION  
ANALYSIS WITH CENSORED DATA.

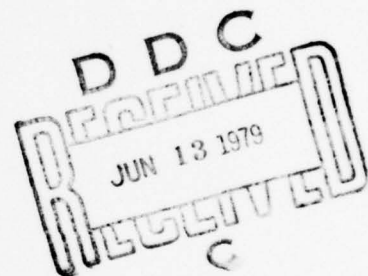
by

10 Josef/Schmee  
Institute of Administration and Management  
Union College and University  
Schenectady, New York 12308

Gerald J./Hahn  
Corporate Research and Development  
General Electric Company  
Schenectady, New York 12308

12 56p.

11 15 April 1979



9 Technical rept.

1 Jun 78-15 Apr 79

15 \*With partial support from the Office of  
Naval Research under Contract No.  
N00014-77-C-0438

This document has been approved  
for public release and sale; its  
distribution is unlimited.

408 725

LB

A SIMPLE METHOD FOR REGRESSION  
ANALYSIS WITH CENSORED DATA

Josef Schmee  
Institute of Administration  
and Management  
Union College  
Schenectady, NY 12308

and

NYS Department of Health  
Albany, NY 12237

Gerald J. Hahn  
Corporate Research  
and Development  
General Electric Company  
Schenectady, NY 12345

ABSTRACT

Problems requiring regression analysis of censored data arise frequently in practice. For example, in accelerated testing one wishes to relate stress and average time to failure from data including unfailed units, i.e., censored observations.

Maximum likelihood is one method for obtaining the desired estimates; in this paper, we propose an alternative approach. An initial least squares fit is obtained treating the censored values as failures. Then, based upon this initial fit, the expected failure time for each censored observation is estimated. These estimates are then used, instead of the censoring times, to obtain a revised least squares fit and new expected failure times are estimated for the censored values. These are then used in a further least squares fit. The procedure is iterated until convergence is achieved. This method is simpler to implement and explain to non-statisticians than maximum likelihood and appears to have good statistical and convergence properties.

The method is illustrated by an example, and some simulation results are described. Variations and areas for further study

also are discussed.

**KEY WORDS:** Censored data, regression analysis, maximum likelihood, accelerated testing, life testing, least squares.

Accession For	
NTIS Grant	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
_____	
Dist	Available or special
A	

## 1. INTRODUCTION

Statistical relationships between variables must often be fitted from data including censored observations. For example, in evaluating material properties, life tests are conducted at accelerated stresses; the resulting data are used to relate stress and average time to failure assuming a model, such as the Arrhenius relationship or the inverse power law. However, at low stress conditions -- the region of greatest interest -- some test units remain unfailed by the time the data must be analyzed, thus resulting in censored observations. Indeed, a test program which is designed to assure that all units fail within the available testing time might require stresses to be accelerated to such an extreme that new failure modes are introduced and the assumed model is no longer valid.

Maximum likelihood is one method for obtaining the desired estimates from the resulting censored data. In practice, however, data from test programs involving censoring are frequently analyzed, not by a professional statistician, but by an engineer or scientist who has had only limited training in statistics. Such persons generally have not been exposed to maximum likelihood methods and may not have easy access to a computer program for performing the complex required calculations. Instead, we have found that censored data are frequently analyzed incorrectly by standard least squares techniques for complete data using one of the following two expedencies to remove the censoring:

- Treating the censored observations as if they were uncensored, e.g., in analyzing product life data, assuming the censoring times to be failure times.
- Ignoring the censored observations altogether.

Both of these procedures lead to biased estimates of the regression line. In particular, they result on the average in overly conservative estimates of product life at design conditions; the first procedure incorrectly assumes that a failure has taken place when, in fact, it has not, while the second approach ignores valid data on unfailed units (often at the conditions of greatest practical interest). These two incorrect procedures also produce biased estimates of the variability around the fitted line, as measured by the residual error variance. Both cases underestimate the values on the average, resulting in too short a confidence interval on the average life at a design condition.

It is therefore desirable to have methods for analyzing censored data which are relatively simple to explain to non-statisticians and to apply by them, in addition to having good statistical properties. One such method, based upon an iterative least squares approach, is proposed here. To apply this method to fit standard relationships for censored data under the usual assumptions, such as a normal or log normal distribution for the random variation, requires only a computer routine for least squares regression analysis and a tabulation or calculation of normal distribution ordinates and areas.

Most of the subsequent discussion will be in the context of product life analysis and thus will be concerned with censoring to the right. Also, we will be principally concerned with the single censoring situation where the censoring times exceed the failure times at each stress, i.e., so-called Type I or time-censoring. As we will indicate later, the basic approach also applies for more general situations.

In summary, the proposed method is: At Step 1 of Iteration 0 obtain an initial least squares fit from the given data treating the censored values as failures. At Step 2 of this iteration use this initial fit to estimate the expected failure time for each unfailed unit, conditional upon its observed censoring time. To obtain a revised least squares fit at Step 1 of Iteration 1, use these estimates of expected failure times for the unfailed units in place of the censoring times. At Step 2 of Iteration 1 estimate from this new fit the new expected running times for the censored observations. At Iteration 2 use these new estimates in a further least squares fit. Iterate this procedure until convergence is achieved.

In Section 2 of this paper, relevant past results are reviewed. Section 3 provides details of the proposed iterative least squares procedure for a simple linear model, normally distributed random variation and censoring to the right. This method is illustrated by a numerical example in Section 4. These analyses should suggest the use of the method for more general situations. In Section 5 the statis-

tical properties of the proposed procedure are evaluated by a simulation analysis and compared with those of the method of maximum likelihood. Generalizations of the results and areas for further study are discussed in Section 6 and 7 respectively, and some concluding remarks are made in Section 8.

## 2. RELATED WORK AND LITERATURE

Hahn and Nelson [3] describe and compare three general methods for analyzing censored life data to estimate relationships between stress and product life:

- Maximum likelihood estimation
- Linear unbiased methods
- Graphical methods.

The method of maximum likelihood is the most general of these approaches and is, perhaps, the one used most frequently by statisticians. However, it requires special computer programs, such as GE STATPAC, see Nelson, Morgan and Capcral [11] which are often not readily available to analysts. Also, maximum likelihood estimates are subject to small sample bias. Graphical methods provide useful supplements to a more formal analysis, but are subjective in nature. The linear estimation methods are computationally simpler than the maximum likelihood methods. However, unlike the procedures we propose in this paper, they are strictly applicable to Type II censoring only. This means, for example, that the information at test conditions where no failures have occurred is ignored in the analysis. These conditions, however, are frequently the ones of greatest practical interest.

Hartley [4] proposes an iterative maximum likelihood approach which bears some similarity to the proposed iterative least squares procedure. In Hartley's paper, the missing observations are replaced by introducing pseudo-frequencies and applying the standard maximum likelihood procedures for complete data.

Further discussion of this method and some extensions are given by Hartley and Hocking [ 5 ], Hughes [ 6 ], and Krane [ 8 ]. In a recent paper, Dempster, Laird and Rubin [ 2 ] describe an algorithm for maximum likelihood estimation; this algorithm can also be applied to the problem considered in this paper (see Section 7).

### 3. DESCRIPTION OF METHOD FOR SIMPLE SITUATION

Assume the standard simple regression model within the region of interest between the stress  $x$  and the average time to failure  $\mu_x$  for some device on life test at that stress, i.e.,

$$\mu_x = \beta_0 + \beta_1 x \quad (1)$$

where, at any stress, (log) time to failure is normally distributed with a constant standard deviation  $\sigma$  and where  $\beta_0$ ,  $\beta_1$  and  $\sigma$  are unknown parameters. One or more units are tested at each of several stresses. The resulting data, after  $c_x$  units of running time at the various stresses, consist of the failure times ( $\leq c_x$ ) on the failed units and the running times  $c_x$  on the unfailed units or "run-outs." At some (usually high stress) conditions, there may be only failures and, at other (usually low stress) conditions, there may be only run-outs.

Let  $c_x$  denote the censoring time for a particular run-out at stress  $x$ . Then, using the well-known properties of the truncated normal distribution (see Johnson and Leone [7] or Nelson [9]), the expected value  $\mu_x^*$  of the failure time for this unit is:

$$\mu_x^* = \mu_x + \sigma f(z) / [1 - F(z)] \quad (2a)$$

where

$$z = (c_x - \mu_x) / \sigma \quad (2b)$$

and  $f(z)$  and  $F(z)$  denote, respectively, the ordinate at  $z$  and the area to the left of  $z$  of a unit normal distribution. Thus, for this situation, the iterative least squares procedure is as follows:

### Iteration 0

Step 1: Fit the linear relationship using standard least squares regression analysis, treating the run-outs as if they failed at their censoring times. Let  $\hat{\beta}_0^{(0)}$ ,  $\hat{\beta}_1^{(0)}$  and  $\hat{\sigma}^{(0)}$  denote the resulting initial estimates of  $\beta_0$ ,  $\beta_1$  and  $\sigma$  respectively.

Step 2: Use the initial regression fit from Step 1 to estimate the unconditional mean time to failure for each run-out. Denote the resulting estimate at stress  $x$  as  $\hat{\mu}_x^{(0)}$ . Then, use the initial estimates  $\hat{\mu}_x^{(0)}$  and  $\hat{\sigma}^{(0)}$ , instead of the unknown values  $\mu_x$  and  $\sigma$ , to estimate the mean failure time  $\hat{\mu}_x^{*(0)}$  for each run-out  $x$  using equations (2a) and (2b).

### Iteration 1

Step 1: Obtain a revised least squares regression fit using the estimated mean failure times  $\hat{\mu}_x^{*(0)}$  obtained in Step 2 of the previous iteration for the run-outs. Let  $\hat{\beta}_0^{(1)}$ ,  $\hat{\beta}_1^{(1)}$  and  $\hat{\sigma}^{(1)}$  denote the new estimates of  $\beta_0$ ,  $\beta_1$  and  $\sigma$ .

Step 2: Repeat Step 2 of the previous iteration using the least squares estimates obtained in Step 1 of the current iteration to re-estimate the mean times to failure for the run-outs.

Subsequent Iterations: Continue the above procedure until convergence is achieved.

As indicated, the preceding iterative least squares procedure requires only a standard computer program for regression analysis and tabulations or calculations of the ordinates and tail areas for a standard normal distribution

or a tabulation of the expected values of a truncated normal variate. In practice, one might wish to mechanize the process by writing a computer routine to perform the iterations. We have developed such a program, using standard approximations for normal distribution ordinates and areas. This program is used for the simulation analysis described in Section 5. It is less complicated than the companion program used in Section 5 for obtaining estimates by the method of maximum likelihood.

The speed of convergence depends on the proportion of censored observation. The iterations in the simulation analysis were stopped after both the slope and intercept estimates agreed to 3 decimal places on two consecutive iterations. In most cases convergence was achieved in fewer than 15 iterations and in very few cases were more than 50 iterations required.

#### 4. NUMERICAL EXAMPLE

Table 1 gives the results of temperature accelerated life tests on electrical insulation in 40 motorettes, originally reported by Crawford [1]. Ten motorettes were tested at each of four temperatures. Testing was terminated at different times at each temperature, resulting in a total of 17 failed units and 23 unfailed ones. The model used to analyze the data assumes that:

- i) for any temperature, the distribution of time to failure is lognormal
- ii) the standard deviation  $\sigma$  of the lognormal time to failure distribution is constant, and
- iii) the mean of the logarithm of the time to failure  $\mu_x$  is a linear function of the reciprocal  $x = 1000/(T+273.2)$  of the absolute temperature  $T$ , that is

$$\mu_x = \beta_0 + \beta_1 x$$

where  $\beta_0$ ,  $\beta_1$  and  $\sigma$  are unknown parameters. The preceding model is often referred to as the Arrhenius relationship.

Nelson and Hahn [10] fitted the given data to the preceding relationship using linear estimation methods, and Hahn and Nelson [3] used graphical and maximum likelihood estimation techniques to fit the same data and compared the results from the various methods. We now fit the data using the iterative least squares method. The procedure is described below and illustrated in Figure 1.

### Iteration 0

Step 1: The censored observations are initially taken as failures, i.e., the 10 unfailed units at 150°C are assumed to be failures at 8064 hours, the 3 unfailed units at 170°C are assumed to be failures at 5448 hours, etc. Then simple least squares is used to fit the assumed relationship. This leads to the following initial parameter estimates:

$$\begin{aligned}\hat{\beta}_0^{(0)} &= -4.9307, \quad \hat{\beta}_1^{(0)} = 3.7471, \\ \hat{\sigma}^{(0)} &= 0.1572\end{aligned}$$

Step 2: From the initial fit, unconditional expected failure times for each of the censored values are estimated. For example, at 220°C, i.e., ( $x = 1000/(220 + 273.2) = 2.03$ ), the mean log time to failure is estimated from the initial fit to be

$$\hat{\mu}^{(0)}_{2.03} = -4.9307 + 3.7471 [1000/(220+273.2)] = 2.6668.$$

Thus, based upon the initial fit, log time to failure at 220°C is estimated to follow a normal distribution with an estimated mean of 2.6668 and an estimated standard deviation of 0.1572.

Censoring at 220°C occurred at 528 hours, i.e.,

$$c_{2.03} = \log(528) = 2.7226$$

If one assumes a log normal distribution for time to failure, the censoring time is equivalent to an estimated normal deviate of

$$z = \frac{c_{2.03} - \hat{\mu}^{(0)}_{2.03}}{\hat{\sigma}^{(0)}} = \frac{2.7226 - 2.6668}{0.1572} = 0.3552$$

The resulting estimate of the expected log time to failure for each of the 5 unfailed units at 220<sup>o</sup>, conditional upon their observed running times of 528 hours, is therefore

$$\hat{\mu}^{*(0)}_{2.03} = \hat{\mu}^{(0)}_{2.03} + \hat{\sigma}^{(0)} \frac{f(z)}{1-F(z)} = 2.6668 + (0.1572) \frac{0.3745}{1-0.6387} = 2.83$$

or 676 hours. Similarly, estimates of the conditional expected failure times for the unfailed units at the other three temperatures are

150<sup>o</sup>C: 4.0384 log hours (or 10,924 hours)

170<sup>o</sup>C: 3.8089 log hours (or 6,440 hours)

190<sup>o</sup>C: 3.3295 log hours (or 2,135 hours)

#### Iteration 1

Step 1: The estimated log times to failure for the censored observations obtained in Step 2 of Iteration 0 are used to fit a new least squares line to the data.

The following parameter estimates result:

$$\hat{\beta}_0^{(1)} = -5.2603, \hat{\beta}_1^{(1)} = 3.9263, \hat{\sigma}^{(1)} = 0.1799$$

Step 2: Using the estimates from Step 1 of this iteration the following revised estimate of the conditional expected time to failure is calculated for each of the censored observations:

150<sup>o</sup>C: 4.09852 log hours (or 12,546 hours)

170<sup>o</sup>C: 3.83972 log hours (or 6,914 hours)

190<sup>o</sup>C: 3.36553 log hours (or 2,320 hours)

220<sup>o</sup>C: 2.85846 log hours (or 722 hours) =  $\hat{\mu}^{*(1)}_{2.03}$

Subsequent Iterations: The preceding steps are repeated until convergence is achieved. This example required 17

iterations. The final parameter estimates are

$$\hat{\beta}_0 = -5.81829, \hat{\beta}_1 = 4.20426, \hat{\sigma} = 0.204322$$

and the final estimates of the expected time to failure for the censored observations are

150°C: 4.17297 log hours (14,893 hours)

170°C: 3.87676 log hours (7,529 hours)

190°C: 3.40090 log hours (2,517 hours)

220°C: 2.87982 log hours (758 hours) =  $\mu^{*(17)}_{2.03}$

Table 2 shows the results obtained at various iterations. In this example results close to the final values, from a practical viewpoint, were obtained after 12 iterations.

The iterative least squares estimates are compared in Table 3 with the maximum likelihood estimates obtained by Hahn and Nelson [3] for the same data. The results from the two methods are quite similar and, except for the residual error standard deviation, agree more closely with each other than with the estimates obtained from graphical and linear unbiased estimation. The latter two methods do not, however, use the test data at 150°C, where none of the 10 test units failed. If one ignores the 150°C data, the iterative least squares estimates and the maximum likelihood estimates were even closer to each other than before and again agreed better with each other than they did with the graphical and linear unbiased estimates (details available from the authors).

## 5. RESULTS OF SIMULATION ANALYSIS

### A. Description of Study

A simulation analysis was used to compare the properties of the iterative least squares procedure with those of the method of maximum likelihood. A linear relationship was assumed between the mean (log) time to failure and the applied (log) stress  $x$ , i.e.,

$$\mu_x = \beta_0 + \beta_1 x$$

and at a given stress, the (log) time to failure was assumed to follow a normal distribution with a constant standard deviation  $\sigma$ , where  $\beta_0$ ,  $\beta_1$  and  $\sigma$  are unknown parameters.

Without any loss of generality,  $\beta_0$  and  $\beta_1$  were set equal to 1 and -1, respectively. Testing was assumed to be conducted uniformly over the interval (0,1), i.e., ranging from the minimum stress of  $x = 0$  (with  $\mu_0 = 1$ ) to the maximum stress of  $x = 1$  (with  $\mu_1 = 0$ ).

Simulation analyses were conducted at each of the 14 different conditions shown in Table 4. The following variables were studied:

1. The censoring probability  $p$  at the minimum stress, i.e., at  $x = 0$ .
2. The ratio  $r$  of the censoring probability at the minimum stress ( $x = 0$ ) to the censoring probability at the maximum stress ( $x = 1$ ).
3. The number of equally spaced test stresses  $k$  over the interval from  $x = 0$  to  $x = 1$ . One test

unit per stress was assumed at 11 of the conditions; multiple test units per stress were evaluated at the other three conditions.

Specification of  $p$  and  $r$ , together with the stated assumptions, defines the censoring time  $c_x$  at each stress and also the common standard deviation  $\sigma$ .

The 14 conditions were as follows:

- The "center" condition of  $p = 0.75$ ,  $r = 5$  and  $k = 10$ ; thus, the censoring probabilities over the 10 stresses ranged from 0.75 at  $x = 0$  to 0.15 at  $x = 1$  (Condition 1).
- The 8 combinations of conditions (in a full factorial arrangement) of  $p = 0.5$  and  $0.95$ ,  $r = 2$  and  $10$  and  $k = 5$  and  $20$  (Conditions 2 through 9).
- The two conditions with  $p = 0.75$  and  $r = 5$  (in both cases) and with  $k = 5$  and  $20$  (Conditions 10 and 11).
- The condition with  $p = 0.75$ ,  $r = 5$  and 3 test units at  $x = 0$  and  $x = 1$  and 4 test units at  $x = 0.5$  (Condition 12).
- The condition with  $p = 0.75$ ,  $r = 5$  and 5 test units at  $x = 0$ , 3 test units at  $x = 0.5$  and 2 test units at  $x = 1$  (Condition 13).
- The condition with  $p = 0.75$ ,  $r = 5$  and 2 test units at  $x = 0$ , 3 test units at  $x = 0.5$  and 5 test units at  $x = 1$  (Condition 14).

Enough simulations were conducted at each of these 14 conditions so as to obtain 1,000 runs (5,000 and

10,000 runs at some conditions) with one or more censored observations and two or more uncensored ones. Table 4 shows for each condition the number of evaluated simulation runs, the average number of censored observations in these runs, and also percent of added unevaluated runs (relative to the number of evaluated runs) at each test condition resulting in 0, k-1 and k censored observations.

In each simulation run, failure times were randomly generated under the previously stated assumed model. The assumed model was fitted using least squares regression analysis based on the complete data (CLS) before censoring. All observations exceeding  $c_x$  were then assumed to be censored at  $c_x$  and new fits were obtained using iterative least squares (ILS) and maximum likelihood (ML).

## B. Summary of Results

### 1. Overall Summaries

Tables 5 and 6 summarize the following results of the simulation analyses:

- The estimated regression intercept and mean at the low stress ( $\beta_0 = \mu_0 = 1$ ) (Tables 5A and 6A).
- The estimated regression slope ( $\beta_1 = -1$ ) (Tables 5B and 6B).
- The estimated mean at the high stress ( $\beta_0 + \beta_1 = \mu_1 = 0$ ) (Table 5C).
- The estimated standardized residual error standard deviation, i.e., the ratio of the calculated residual error standard deviation

to the true residual error standard deviation ( $\sigma' = 1$ ) (Tables 5D and 6C).

Each part of Table 5 provides a comparison, for each test condition, of (1) the average, (2) the standard deviation, (3) the root mean square error and (4) the minimum and maximum values of the desired estimates using the CLS, ILS and ML methods. This table is based upon those simulation runs which resulted in one or more censored observations and two or more uncensored ones. The simulations which did not result in any censored observations were excluded because, in these cases, CLS, ILS and ML provide identical results (except that the ML estimate of the residual error standard deviation is smaller than the CLS and ILS estimates by a factor of  $[(n-2)/n]^{1/2}$ ). Also, those simulations with all censored observations or with only a single uncensored observation were excluded because no ML estimates can be obtained for them. As a result of these exclusions, the resulting (conditional) CLS estimates of  $\beta_0$  and  $\beta_1$  are no longer unbiased.

Table 6 provides, for each test condition, (1) paired comparisons of the average difference, (2) standard deviation of differences (3) minimum and maximum differences in estimates for ILS - ML, ILS - CLS, and ML - CLS with regard

to  $\beta_0$  and  $\beta_1$  and similar statistics for the ratios with regard to  $\sigma^2$  and

(4) evaluations of Pitman closeness (see Pitman [13]), i.e., the percentage of simulation runs for which:

- The ILS estimates are closer to the true parameter value than are the ML estimates.
- The ILS estimates are closer to the true parameter value than are the CLS estimates.
- The ML estimates are closer to the true parameter value than are the CLS estimates.

## 2. Specialized Summaries

Table 7 provides detailed two-way tabulations showing the estimates of the regression intercept and, equivalently, the mean at the low stress ( $\beta_0 = 1$ ) from the 5000 simulations conducted at Condition 1 for the following:

- Iterative least squares versus maximum likelihood (Table 7A).
- Iterative least squares versus complete least squares (before censoring) (Table 7B).
- Maximum likelihood versus complete least squares (before censoring) (Table 7C).

Although not given, similar detailed comparisons have been obtained for the estimates of the slope coefficient and of the residual error standard deviation, and also for each of the other 13 conditions.

### C. Discussion of Results

Inspection of Tables 5, 6 and 7 indicates that, for most of the conditions considered, the iterative least squares method and the method of maximum likelihood yield very similar results (see especially Table 7A). A notable exception was Condition 6 (discussed further below).

As one might expect, in some cases both ILS and ML gave biased estimates of appreciable magnitude at some or all of the 5 conditions where there was a large number of unevaluated simulations due to obtaining 0,  $k-1$  or  $k$  censored observations (Conditions 2,3,4,5 and 10). In these cases, the evaluated CLS estimates are also biased, although sometimes the magnitude of the bias is smaller. In addition to these 5 conditions, estimated biases exceeding 10% resulted in

- Estimating  $\beta_0$  and  $\beta_1$  using both ILS and ML at two further conditions.
- Estimating  $\beta_0 + \beta_1$  (i.e., the mean at the high stress condition) using ILS at one further condition.
- Estimating  $\sigma'$  using ILS at seven further conditions and using ML at four further test conditions.

The estimated biases with regard to  $\beta_0$ ,  $\beta_1$  and  $\beta_0 + \beta_1$  were positive at some conditions and negative at others. In evaluating  $\sigma'$ , however, underestimates were obtained, on the average, at almost all conditions for both ILS and ML.

An overall comparison of the performance of the iterative least squares and maximum likelihood estimation methods at the 14 test conditions in the simulation analysis is shown in Table 8. This tabulation shows that there were

- An approximately equal number of conditions where the average ILS estimate is closer to the true parameter value than is the average ML estimate.
- A larger number of conditions at which the root mean square error is smaller for the ILS estimate than for the ML estimate, especially with regard to  $\beta_0$  and  $\beta_1$ .
- A larger number of conditions at which the ILS estimate has more simulation runs closer to the true parameter value than does the ML estimate with regard to  $\beta_0$ ,  $\beta_1$  and  $\sigma'$ . In estimating  $\sigma'$ , however, the average percentage of runs (averaged over the 14 conditions, weighting each condition equally) at which the ILS estimate is closer to  $\sigma'$  was essentially equal to that for which the ML estimate was closer.

On balance, then, the proposed iterative least squares procedure gave results which are at least competitive with those obtained by maximum likelihood. Iterative least squares did poorest relative to maximum likelihood at Condition 6, although maximum likelihood results at that condition were also quite poor. The censoring probability at this condition ranges from 0.95 at the low stress to 0.475 at the high stress and, on

the average, about three-quarters of the 20 observations are censored, the highest proportion among the 14 test conditions. The large number of run-outs, especially at or near the low stress, (1) greatly bias the initial fit and this is not fully compensated for in the later iterations and (2) lead to an appreciable underestimate of the residual error standard.

An underestimate of the residual error standard deviation estimated by the iterative least squares method versus that from maximum likelihood also was found at two other conditions (Conditions 8 and 14) at which there are multiple censored observations at the same or neighboring stresses. (This might also have been the case in the motorette example of Section 3.) Since the censored observations at the same (or similar) stresses have the same (or similar) estimates of the expected value, the variability among them is small and thus contributes little to the overall estimate of variability. Some possible remedies for this are suggested in Section 7.

In conclusion, we note that the relatively good results of iterative least squares when compared with maximum likelihood may be due to the small sample bias in the maximum likelihood estimates. An assessment of the magnitude of this bias is a by-product of our simulation analysis.

D. Comment on Computational Requirements

A major reason for proposing the iterative least squares approach is that the calculations are less than those required for maximum likelihood estimation. For example, the iterative least squares calculations can be programmed with fewer than 50 FORTRAN statements, whereas the maximum likelihood calculations using a Powell optimization routine require approximately 300 FORTRAN statements. The average number of function evaluations for the 14 test conditions in the simulation analysis was about 15. For problems of this type, maximum likelihood usually requires at least 50 function evaluations. It is possible, however, that some of the more recently suggested maximum likelihood algorithms (see, for example, Dempster, Laird and Rubin [2]) reduce the required calculational effort for this method.

## 6. GENERALIZATION OF RESULTS

Most of the preceding discussion has dealt with the special situation where:

- There is a simple linear relationship (between stress and time to failure).
- The random variation follows a (log) normal distribution.
- Censoring is to the right, i.e., the censored values equal or exceed the observed values and occur at the same value (or time) for a particular condition (or stress).

The iterative least squares procedure is not, however, limited to this specialized situation.

### More general regression relationships

The proposed method handles censoring of the dependent variable. It is not limited to any special form of relationship and can be applied to multiple regression or, more generally, to any situation for which least squares procedures are used.

### Other distributions for random variation

Let  $f(y_x)$  and  $F(y_x)$  denote the probability density function and cumulative distribution function, respectively, of the random variable  $y$  at the condition  $x$  and let censoring to the right occur at the value  $c_x$  of  $y$ . Also let  $y_x^*$  denote the value which would be observed for a censored observation if it were observable, e.g., the time at which a run-out would fail if there were no censoring. Then the conditional probability density function of an observation censored to the right, i.e., one which is known to exceed the censoring

value  $c_x$ , is that of the random variable  $y_x$  truncated to the right at  $c_x$ . Therefore

$$f(y_x^*) = f(y_x) / [1 - F(c_x)] , y_x^* > c_x$$

and the conditional expected value of the censored observation (i.e. its expected failure time) is

$$\mu_x^* = \int_{c_x}^{\infty} y_x f(y_x) dy_x / [1 - F(c_x)] . \quad (3)$$

Thus, the proposed method requires evaluation of the conditional expectation given in equation (3). If the random variation is normally distributed this results in equations (2a) and (2b). Similar results can be obtained for other distributions for the random variation.

#### Other censoring schemes

In practice, one may encounter situations involving censoring other than to the right. For example, in a life test, censoring may occur at random times due to accident or breakdown of equipment. Also, censoring other than to the right can occur when failures are discovered only at periodic inspections.

The iterative least squares concept, just as the method of maximum likelihood, applies to such more general censoring schemes. However, when the censoring occurs early, it might be better in the initial least squares fit to ignore the censored observations instead of treating them as failures. Also, a procedure which assigns varying lesser weights to the censored observations than to the uncensored ones (see next section) might be especially appropriate for more general censoring schemes.

## 7. AREAS FOR FURTHER STUDY

A theoretically based study of the properties of the proposed iterative least squares approach is warranted. In addition, some specific areas for further consideration are reviewed below.

### Assigning less weight to censored observations

The informational content of the censored observations is less than that of the uncensored ones. For example, in censoring to the right, the earlier the censoring occurs, the less is the information provided by the censored observation. This might be taken into consideration in the iterative least squares analysis by assigning less weight to the censored observations than to the uncensored ones and by varying these weights depending on the censoring point. The following is one possible weighting scheme for the censored observation  $y_x^*$  at the condition  $x$ :

$$w_1(x) = 1 - [\sigma^2(y_x^*)/\sigma^2]$$

where  $\sigma$  denotes the (constant) residual error standard deviation and  $\sigma(y_x^*)$  denotes the standard deviation of the truncated distribution of an observation known to exceed  $c_x$ . For example, in a product life analysis with censoring to the right,  $\sigma$  denotes the standard deviation of the distribution of time to failure at a constant stress and  $\sigma(y_x^*)$  denotes the standard deviation of the distribution of the time to failure of a unit at stress  $x$  which is known to be unfailed at time  $c_x$ . In particular, for normally distributed random variation, it can be shown (see Nelson [9]) that, using the

notation of Section 3,

$$\sigma(y_x^*) = \sigma \{ 1 + [zf(z) / [1 - F(z)]] - [f(z)/(1 - F(z))]^2 \}^{1/2}$$

where, as before,  $z = (c_x - \mu_x)/\sigma$ .

This procedure assigns a weight of 1 to all uncensored observations and, for normally distributed random variation, leads to a weight which approaches 0 with decreasing  $c_x$  and a weight of 1 with increasing  $c_x$ . This scheme may, however, be inappropriate when the random variation is not normally distributed. For example, if the time to failure distribution is exponential, all the censored observations would receive a weight of 0. For a Weibull distribution with a shape parameter less than 1, censored observations would receive negative weights.

An alternative simple scheme is to weight the censored observation proportional to the associated cumulative (failure) probability at the censoring point, i.e., to assign to an observation censored to the right of  $c_x$  the weight

$$w_2(x) = F(c_x)$$

and to give all uncensored observations a weight of 1.

In practice, the required distribution ordinates and cumulative probabilities are, of course, unknown. In their place one would use the estimates obtained in the most recent iteration. For the initial iteration, some working rule needs to be developed to obtain an appropriate (initial) weight between 0 and 1 for the censored observations.

The reason for using weighted iterative least squares estimation is the expectation that it will lead to better

performance than the proposed unweighted estimator. The weighting procedure, however, complicates the method and thus compromises our claim of simplicity. Despite this, a computer program based upon a weighted iterative least squares approach is still likely to be appreciably simpler to develop, use and explain than one based upon maximum likelihood.

Development of statistical inference procedures and modification of standard regression programs

The preceding discussion has been concerned with obtaining point estimates using the proposed iterative least squares procedure. Frequently, statistical intervals and hypothesis tests are also desired. With the method of maximum likelihood, this is accomplished using asymptotic theory. Such inferences tend to be unconservative for small sample size situations.

Using an iterative least squares approach, one can obtain unconservative inferences by applying the usual least squares inference procedures to the results of the final iteration and treating the censored observations as if they were uncensored. In addition, it is possible to obtain conservative confidence intervals and prediction intervals by ignoring the censored observations altogether in obtaining the length of the desired interval (but not in obtaining the point estimate around which the interval is constructed) after, perhaps, adjusting for bias in the estimate of the residual error standard deviation.

A compromise between such overly unconservative and overly conservative approximations would be to calculate the bounds, as well as the original estimates, using a scheme which assigns varying weights to the censored observations (see above discussion).

Along similar lines, consideration might be given to appropriate ways of adapting standard computer programs for regression analysis to handle residual plots with censored data and for performing various standard analysis of variance tests to assess the significance and adequacy of a fitted model with censored data. In particular, a referee has suggested a lack of fit test based on censored versus non-censored errors as a natural extension of our procedures and as a useful method of evaluating the lack of fit of the assumed equation (see also Nelson [12]).

#### Specialization to a single condition

The iterative least squares procedure is not limited to situations dealing with a regression relationship. It can similarly, and more simply, be applied when data at a single condition are available and estimates are desired only for the parameters of the distribution for the random variation at that condition. We have been concerned with the more general regression situation because (i) this is the type of practical problem which we have encountered most frequently, and (ii) special simplified procedures are already available for the single condition situation. However, this situation might be useful for studying various generalizations and extensions, such as the properties of iterative least squares under different censoring

schemes or the effect of different start-up or weighting procedures. The simpler single condition situation might also be appropriate for a theoretical study of the properties of the proposed procedures.

#### Use of results from early iterations

The number of iterations to convergence is important for those who use an iterative least squares approach to analyze censored data via a standard least squares computer routine or, for example, a hand calculator. It is also of interest to determine how close one can come to the final convergence values after a small number, such as 3, 4, or 5, iterations. In this regard, procedures to speed convergence also merit consideration.

#### Bias adjustment procedures

The results of the simulation analysis indicate bias at some of the conditions in the iterative least squares estimates, as well as in the maximum likelihood estimates, especially in the estimate of the residual error standard deviation. The simulation results might themselves be used to obtain appropriate bias correction factors.

Procedures for correcting for the underestimate of the residual error standard deviation within the iterative scheme also warrant consideration. An upward adjustment factor might be arrived at based upon the proportion of censored observations in the data. One possibility is to reduce the degrees of freedom in calculating the residual error standard deviation. In a private communication, R. Regal has suggested to us an alternative scheme for calculating the residual error standard deviation based upon the

EM-algorithm discussed in a paper by Dempster, Laird and Rubin [2]. The EM procedure can be viewed as a generalization of the method proposed by Hartley [4] and by other authors referenced in [2]. It uses the same expression for the slope and intercept estimates as our proposed method.

## 8. CONCLUSION

This paper proposes an alternative scheme for analyzing censored data which might be more attractive to some practitioners than existing methods. This is because the proposed procedure uses well-known least squares techniques, requires less computation than maximum likelihood, and should be easier to explain to non-statisticians.

A simulation analysis indicates that the proposed method gives statistically reasonable results which, for the situations considered in the evaluations, are at least competitive with those from maximum likelihood. Although the simulations were limited to cases where censoring is to the right and where the censored values exceed the uncensored ones, this is the situation which is frequently encountered in the analysis of accelerated life data and many other practical applications.

As indicated in Sections 6 and 7, various generalizations of these results are possible. To extend and refine the results obtained to date, further work is warranted.

#### ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. Paul Feder of Ohio State University, to Dr. Wayne Nelson of General Electric Corporate Research and Development, and to Dr. Ron Regal of the State University of New York in Albany for their useful comments on an earlier draft of this paper and to the referees for their constructive suggestions.

Dr. Josef Schmee received support from the Office of Naval Research under Contract Number N00014-77-C-0438, for which Prof. L.A. Aroian of Union College is project director.

# REFERENCES

- [ 1 ] CRAWFORD, D.E. (1970). Analysis of incomplete life test data on motorettes. Insulation/Circuits 16, 10, 43-48.
- [ 2 ] DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977). Maximum likelihood via the EM algorithm. J. Royal Statist. Soc., B-39, 1-22.
- [ 3 ] HAHN, G.J. and NELSON, W.B. (1974). A comparison of methods for analyzing censored life data to estimate relationships between stress and product life. IEEE Transactions on Reliability, R-23, 2-11.
- [ 4 ] HARTLEY, H.O. (1958). Maximum likelihood estimation from incomplete data. Biometrics, 14, 174-194.
- [ 5 ] HARTLEY, H.O. and HOCKING, R.R. (1971). The analysis of incomplete data. Biometrics, 27, 783-823.
- [ 6 ] HUGHES, E.J. (1962). Maximum likelihood estimation of distribution parameters from incomplete data. Unpublished Ph.D. dissertation, Iowa State University.
- [ 7 ] JOHNSON, N.L. and LEONE, F.C. (1976). Statistics and Experimental Design in the Physical and Engineering Sciences, Vol. 2, revised edition. New York: John Wiley and Sons.
- [ 8 ] KRANE, S.A. (1957). Maximum likelihood estimation from incomplete data for continuous distributions. Unpublished M.S. thesis, Iowa State University.
- [ 9 ] NELSON, W.B. (1967). The truncated normal distribution - with application to component sorting. Industrial Quality Control, 24, 261-271.

- [10] NELSON, W. and HAHN, G.J. (1972 and 1973). Linear estimation of a regression relationship from censored data - Part I: Simple methods and their application. Technometrics, 14, 247-269; Part II: Best linear unbiased estimation and theory. Technometrics, 15, 133-150.
- [11] NELSON, W.B., MORGAN, C.B. and CAPORAL, P. (1978). STATPAC simplified -- A short introduction to how to run STATPAC, a general statistical package for data analysis. General Electric Company TIS Report 78CRD276. (Available from GE CRD Distribution, P.O. Box 43, Bldg. 5, Room 237, Schenectady, NY 12301.)
- [12] NELSON, W.B. (1973). Analysis of residuals from censored data. Technometrics, 15, 697-715.
- [13] PITMAN, E.J.E. (1937). The "closest" estimates of statistical parameters. Proc. Cambridge Philosophical Soc. 32, 212-222.



TABLE 1

## INSULATION LIFE IN HOURS AT VARIOUS TEST TEMPERATURES

Test Temperature			
<u>150°C<sup>1)</sup></u>	<u>170°C<sup>2)</sup></u>	<u>190°C<sup>3)</sup></u>	<u>220°C<sup>4)</sup></u>
	1764	408	408
	2772	408	408
	3444	1344	504
	3542	1344	504
	3780	1440	504
	4860		
	5196		

- 1) All 10 motorettes at 150°C still on test without failure at 8064 hours.
- 2) 3 motorettes at 170°C still on test without failure at 5448 hours.
- 3) 5 motorettes at 190°C still on test without failure at 1680 hours.
- 4) 5 motorettes at 220°C still on test without failure at 528 hours.

TABLE 2  
ITERATIVE LEAST SQUARES ESTIMATES AFTER VARIOUS NUMBER OF ITERATIONS FOR  
INSULATION LIFE DATA EXAMPLE

Iteration $i$	Estimate of log failure time $\mu_x$ used in iteration $i$ for censored observations at			Resulting Estimates of		
	150°C	170°C	$\hat{\mu}_x(i-1)$ 190°C	Intercept Coefficient $\hat{\beta}_0(i)$	Slope Coefficient $\hat{\beta}_1(i)$	Standard Deviation $\hat{\sigma}(i)$
0	3.90655*	3.7362*	3.22531*	-4.9307	3.7471	0.157209
1	4.03838	3.80888	3.32946	-5.2603	3.92631	0.179850
2	4.09852	3.83972	3.36553	-5.48598	4.04036	0.191041
3	4.13064	3.85598	3.38197	-5.62322	4.10845	0.196873
4	4.14861	3.86489	3.39034	-5.70428	4.14838	0.200065
5	4.15886	3.86991	3.39489	-5.75172	4.17167	0.201862
12	4.17266	3.87660	3.40076	-5.81688	4.20356	0.204265
17	4.17297	3.87676	3.40090	-5.81829	4.20426	0.204322

\* Censoring time in test program.

TABLE 3  
COMPARISON OF RESULTS FROM ITERATIVE LEAST SQUARES  
AND MAXIMUM LIKELIHOOD FITS FOR INSULATION LIFE DATA EXAMPLE

	Estimates from	
	Iterative Least Squares	Maximum Likelihood Method
Intercept Coefficient ( $\beta_0$ )	-5.818	-6.027 (+1.489*)
Slope Coefficient ( $\beta_1$ )	4.204	4.314 (+0.684*)
Log Standard Deviation ( $\sigma$ )	0.204	0.259
Median Life in Hours at:		
220°C	508	530
190°C	1812	1940
170°C	4654	5080
150°C	13060	14680
130°C***	40638	47000
		(27500 to 79500**)
10% Point of Life Distribution at:		
130°C***	22230	21900

\*Bounds of approximate 90% confidence interval

\*\*Approximate 90% confidence interval

\*\*\*Extrapolated beyond range of data

TABLE 4

## SUMMARY OF TEST CONDITIONS IN SIMULATION ANALYSIS

Test Condition	Censoring Probability at $x=0(p)$	Ratio of Censoring Probability at $x=0$ to that at $x=l(r)$	Number of Test Stresses $(k)^*$	Number of Evaluated Simulations	Average Number of Censored Observations in Evaluated Simulations	Added Unevaluated Simulations in %**			
						with Following Number			
						Censored Observations	0	k-1	k
1	0.75	5.	10	5000	4.4	0.12	0.12	0.10	0.00
2	0.95	2.	5	10000	2.8	0.07	113.60	54.40	168.07
3	0.50	2.	5	5000	1.9	11.28	6.44	0.60	18.32
4	0.95	10.	5	5000	2.4	0.28	19.22	1.68	21.18
5	0.50	10.	5	5000	1.5	28.44	0.76	0.08	29.28
6	0.95	2.	20	1000	15.1	0.00	1.80	0.30	2.10
7	0.50	2.	20	1000	7.4	0.00	0.00	0.00	0.00
8	0.95	10.	20	1000	11.1	0.00	0.00	0.00	0.00
9	0.50	10.	20	1000	4.7	0.50	0.00	0.00	0.50
10	0.75	5.	5	1000	2.1	4.20	9.40	0.80	14.40
11	0.75	5.	20	1000	8.8	0.00	0.00	0.00	0.00
12	0.75	5.	3 (3,4,3)	1000	4.4	0.20	0.20	0.00	0.40
13	0.75	5.	3 (5,3,2)	1000	5.3	0.00	0.70	0.00	0.70
14	0.75	5.	3 (2,3,5)	1000	3.5	0.08	0.00	0.00	0.08

\*One test per stress for conditions 1 to 11; number of tests at stresses of 0, 0.5 and 1.0 shown in parentheses for conditions 12 to 14.

\*\*Relative to evaluated simulations.

TABLE 5A

OVERALL RESULTS OF SIMULATION ANALYSIS: ESTIMATED REGRESSION INTERCEPT AND MEAN AT LOW STRESS ( $\beta_0 = 1$ )

Test Condition	Average		Standard Deviation		Root Mean Square Error		Minimum and Maximum	
	CLS	ILS	CLS	ILS	CLS	ILS	CLS	ML
1	0.998	1.014	0.344	0.460	0.344	0.461	-0.43 2.12	-0.41 10.54
2	0.906	0.603	0.484	0.564	0.493	0.690	-0.78 2.76	-0.87 5.82
3	1.066	1.303	1.092	1.417	1.094	1.449	-2.98 4.80	-3.21 13.88
4	0.987	0.980	0.268	0.472	0.268	0.472	0.05 1.86	0.07 4.28
5	1.112	1.267	0.426	0.605	0.442	0.661	-0.57 3.00	-0.63 6.41
6	1.010	0.636	0.259	0.439	0.260	0.571	0.24 1.76	-0.02 9.68
7	0.996	0.985	0.638	0.690	0.638	0.690	-1.54 2.90	-1.59 4.38
8	1.008	0.935	0.147	0.239	0.147	0.248	0.53 1.45	0.52 3.20
9	0.993	1.009	0.258	0.285	0.258	0.285	0.14 1.77	0.11 2.24
10	1.014	1.095	0.415	0.642	0.415	0.649	-0.14 2.59	-0.25 5.94
11	1.007	0.975	0.247	0.284	0.247	0.285	0.12 1.88	0.11 2.56
12	1.001	1.008	0.299	0.423	0.299	0.423	0.06 2.04	0.05 12.54
13	1.000	0.925	0.250	0.307	0.250	0.316	0.11 1.72	0.15 7.13
14	1.008	1.134	0.363	0.657	0.363	0.671	-0.19 2.33	-0.15 13.64

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood.

TABLE 5B

OVERALL RESULTS OF SIMULATION ANALYSIS: ESTIMATED REGRESSION SLOPE ( $\beta_1 = -1$ )

Test Condition	Average		Standard Deviation		Root Mean Square Error		Minimum and Maximum		
	CLS	ILS	CLS	ILS	CLS	ILS	CLS	ILS	ML
1	-1.001	-1.021	0.581	0.711	0.581	0.806	-2.92 1.34	-5.78 1.45	-11.20 1.71
2	-1.210	-0.927	0.769	0.883	0.798	0.919	-4.09 1.85	-7.47 2.86	-7.39 3.28
3	-1.101	-1.236	1.929	2.413	1.932	2.318	-6.96 4.94	-17.60 11.98	-18.45 12.62
4	-1.037	-1.031	0.434	0.679	0.435	0.717	-2.66 0.81	-5.02 0.62	-5.29 0.73
5	-1.111	-1.260	0.782	0.979	0.790	1.012	-3.95 1.79	-6.96 2.29	-7.28 2.15
6	-1.023	-0.782	0.439	0.628	0.439	1.151	-2.29 0.49	-4.42 0.36	-10.40 0.52
7	-.980	-.976	1.105	1.191	1.105	1.191	-4.55 2.91	-5.04 3.34	-5.29 3.34
8	-1.003	-.932	0.250	0.354	0.250	0.446	-1.77 -0.28	-3.06 -0.15	-3.79 -0.20
9	-.991	-1.008	0.438	0.471	0.438	0.478	-2.24 0.37	-2.61 0.59	-2.64 0.59
10	-1.059	-1.124	0.704	0.986	0.706	1.004	-3.69 1.15	-7.03 1.51	-7.40 1.43
11	-1.008	-0.980	0.433	0.484	0.433	0.485	-2.59 0.54	-2.68 0.51	-3.18 0.52
12	-1.008	-1.016	0.502	0.626	0.502	0.952	-2.87 0.64	-4.20 0.89	-12.50 0.90
13	-1.022	-0.945	0.478	0.563	0.478	0.566	-2.35 0.51	-3.68 0.45	-7.84 2.91
14	-1.026	-1.156	0.480	0.753	0.486	0.769	-2.85 1.05	-6.63 1.08	-13.11 1.10

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood.

TABLE 5C

OVERALL RESULTS OF SIMULATION ANALYSIS: ESTIMATED MEAN AT HIGH STRESS ( $\beta_0 + \beta_1 = 0$ )

Test Condition	Average		Standard Deviation		Root Mean Square Error		Minimum and Maximum	
	CLS	ILS	CLS	ILS	CLS	ILS	CLS	ML
1	-0.003	-0.007	-0.005	0.361	0.373	0.361	-1.25	-1.52
							1.24	1.45
2	-0.304	-0.324	-0.341	0.375	0.364	0.496	-1.85	-2.02
							1.59	2.11
3	-0.036	0.067	-0.022	1.343	1.257	1.345	-4.28	-4.49
							3.97	9.56
4	-0.049	-0.051	-0.055	0.275	0.270	0.279	-1.10	-1.42
							0.86	0.92
5	0.007	0.007	0.001	0.541	0.514	0.541	-1.66	-1.66
							1.86	2.60
6	-0.013	-0.146	-0.047	0.244	0.345	0.268	-0.68	-0.92
							0.77	0.64
7	0.015	0.009	0.025	0.673	0.696	0.673	-2.11	-1.97
							1.92	2.25
8	0.005	0.002	-0.005	0.153	0.163	0.153	-0.49	-0.55
							0.42	0.43
9	0.001	0.001	0.001	0.268	0.268	0.268	-0.84	-0.87
							0.94	0.97
10	0.045	0.029	0.046	0.487	0.461	0.488	-1.41	-1.59
							1.33	1.74
11	0.001	-0.005	0.002	0.268	0.279	0.268	-0.90	-0.88
							0.93	0.95
12	-0.007	-0.008	-0.004	0.331	0.340	0.331	-1.18	-1.08
							0.83	1.31
13	-0.022	-0.019	-0.009	0.370	0.406	0.370	-1.12	-1.20
							1.07	1.29
14	0.018	-0.022	0.020	0.267	0.272	0.268	-0.74	-0.83
							0.87	1.02

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood.

TABLE 5D

OVERALL RESULTS OF SIMULATION ANALYSIS: ESTIMATED STANDARDIZED RESIDUAL ERROR STANDARD DEVIATION ( $\sigma' = 1$ )

Test Condition	Average		Standard Deviation		Root Mean Square Error		Minimum and Maximum	
	CLS	ILS	CLS	ILS	CLS	ILS	CLS	ML
1	0.966	0.843	0.863	0.305	0.247	0.343	0.25 2.32	0.04 4.31
2	1.005	0.557	0.468	0.466	0.409	0.643	0.05 2.87	0.02 2.25
3	0.972	0.979	0.740	0.561	0.391	0.562	0.04 3.16	0.002 3.76
4	0.922	0.605	0.520	0.498	0.399	0.636	0.04 2.36	0.005 3.21
5	0.982	1.009	0.765	0.544	0.392	0.544	0.02 2.43	0.003 4.18
6	0.991	0.494	0.901	0.212	0.169	0.549	0.52 1.59	0.05 1.51
7	0.985	0.910	0.942	0.190	0.169	0.210	0.52 1.55	0.34 1.90
8	0.985	0.725	0.927	0.193	0.166	0.336	0.48 1.56	0.18 1.39
9	1.001	0.977	0.961	0.188	0.167	0.189	0.55 1.63	0.49 1.81
10	0.933	0.824	0.647	0.554	0.407	0.581	0.09 2.39	0.003 3.11
11	0.979	0.847	0.931	0.192	0.165	0.245	0.47 1.621	0.32 1.53
12	0.976	0.835	0.864	0.292	0.248	0.336	0.31 1.82	0.15 1.86
13	0.962	0.730	0.811	0.306	0.242	0.408	0.35 1.84	0.005 1.94
14	0.956	0.874	0.865	0.284	0.251	0.311	0.32 2.28	0.19 1.87

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood

TABLE 6A

PAIRED COMPARISON RESULTS OF SIMULATION ANALYSIS:  
ESTIMATED REGRESSION INTERCEPT AND MEAN AT LOW STRESS

Test Condition	Average Differences			Standard Deviation of Differences			Minimum and Maximum Differences			Percent Simulations Closer to True Value		
	ILS -ML	ILS -CLS	ML -CLS	ILS -ML	ILS -CLS	ML -CLS	ILS -ML	ILS -CLS	ML -CLS	ILS vs ML	ILS vs CLS	ML vs CLS
1	-0.046	0.015	0.061	0.174	0.324	0.444	-5.97 0.05	-0.89 3.34	-0.86 9.31	83.9	43.8	39.4
2	0.026	-0.304	-0.329	0.085	0.560	0.578	-0.37 0.39	-2.06 4.30	-2.09 4.67	73.8	30.3	26.1
3	0.169	0.237	0.068	0.203	0.877	0.847	-0.76 1.26	-2.99 9.53	-3.07 10.29	44.0	45.0	48.6
4	0.007	-0.007	-0.014	0.060	0.423	0.454	-0.25 0.26	-1.17 2.79	-1.16 3.04	79.6	34.8	29.5
5	0.076	0.149	0.073	0.081	0.402	0.399	-0.36 0.50	-1.14 4.06	-1.17 4.35	39.8	44.4	47.8
6	-0.547	-0.375	0.172	0.642	0.408	0.937	-6.69 -0.03	-1.27 2.40	-1.14 8.83	31.7	17.0	24.1
7	-0.041	-0.011	0.030	0.096	0.308	0.357	-0.90 0.02	-1.11 1.39	-1.075 2.180	69.8	46.4	43.6
8	-0.130	-0.073	0.057	0.131	0.199	0.298	-1.25 -0.00	-0.57 1.48	-0.52 2.13	42.6	31.5	27.9
9	0.000	0.016	0.016	0.016	0.113	0.121	-0.22 0.01	-0.33 0.51	-0.34 0.73	68.4	45.3	44.4
10	0.060	0.080	0.021	0.092	0.532	0.558	-0.34 0.41	-1.04 4.07	-1.06 4.41	69.8	45.1	42.1
11	-0.058	-0.032	0.026	0.076	0.178	0.225	-0.63 0.01	-0.50 0.76	-0.48 1.24	56.3	42.8	39.3
12	-0.091	0.007	0.098	0.519	0.310	0.744	-8.44 0.04	-0.75 2.62	-0.77 11.06	86.4	40.2	35.1
13	-0.101	-0.075	0.026	0.329	0.250	0.504	-4.46 0.03	-0.82 1.68	-0.81 5.83	67.7	40.9	34.0
14	-0.192	0.126	0.319	0.903	0.516	1.350	-9.46 0.06	-0.76 4.07	-0.76 12.65	80.0	43.3	41.1

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood

TABLE 6B

PAIRED COMPARISON RESULTS OF SIMULATION ANALYSIS:  
ESTIMATED REGRESSION SLOPE

Test Condition	Average Differences			Standard Deviation of Differences			Minimum and Maximum Differences			Percent Simulations Closer to True Value		
	ILS -ML	ILS -CLS	ML -CLS	ILS -ML	ILS -CLS	ML -CLS	ILS -ML	ILS -CLS	ML -CLS	ILS vs ML	ILS vs CLS	ML vs CLS
1	0.043	-0.020	-0.064	0.177	0.411	0.529	-0.57 6.31	-3.98 1.44	-10.29 1.57	64.8	43.7	42.0
2	-0.009	0.283	0.292	0.082	0.676	0.701	-0.49 0.41	-5.06 2.45	-5.47 2.45	67.6	48.0	46.0
3	-0.079	-0.134	0.055	0.250	1.148	1.119	-1.33 1.19	-11.42 8.00	-12.22 8.64	27.9	39.4	46.0
4	-0.003	0.006	0.009	0.628	0.511	0.547	-0.23 0.28	-3.19 1.36	-3.46 1.40	73.0	35.5	33.0
5	-0.068	-0.149	-0.080	0.097	0.491	0.488	-0.59 0.40	-4.83 1.41	-5.15 1.41	34.7	39.9	45.2
6	0.448	0.242	-0.206	0.645	0.509	1.031	-0.28 6.94	-3.10 1.41	-9.74 1.41	45.4	32.4	28.3
7	0.025	0.005	0.021	0.092	0.435	0.482	-0.33 0.93	-1.93 1.47	-2.70 1.54	61.0	45.0	43.9
8	0.137	-0.071	-0.066	0.136	0.259	0.360	0.00 1.34	-1.85 0.64	-2.58 0.62	43.1	35.0	33.7
9	0.000	-0.017	-0.017	0.018	0.150	0.158	-0.01 0.24	-0.60 0.45	-0.81 0.46	59.2	45.5	44.7
10	0.043	-0.065	-0.022	0.098	0.647	0.676	-0.41 0.37	-4.78 1.51	-5.15 1.51	54.8	39.5	39.7
11	-0.058	0.028	-0.028	0.076	0.237	0.283	-0.63 0.01	-1.06 0.78	-1.59 0.77	56.3	44.3	42.5
12	0.087	-0.008	-0.095	0.510	0.368	0.776	-0.22 8.29	-2.91 0.96	-11.20 1.17	69.8	39.9	38.5
13	0.091	0.077	-0.014	0.337	0.311	0.562	-2.65 4.47	-2.06 0.98	-6.21 2.96	62.2	41.6	38.7
14	0.190	-0.131	0.321	0.888	0.547	1.360	-0.06 9.05	-4.19 0.89	-12.25 0.84	66.7	42.1	40.5

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood

TABLE 6C

PAIRED COMPARISON RESULTS OF SIMULATION ANALYSIS:  
ESTIMATED STANDARDIZED RESIDUAL ERROR STANDARD DEVIATION

Test Condition	Average Ratio			Standard Deviation of Ratios			Minimum and Maximum Ratios			Percent Simulations Closer to True Value		
	ILS /ML	ILS /CLS	ML /CLS	ILS /ML	ILS /CLS	ML /CLS	ILS /ML	ILS /CLS	ML /CLS	ILS vs ML	ILS vs CLS	ML vs CLS
1	0.987	0.868	0.891	0.121	0.204	0.238	0.48 1.92	0.06 1.56	0.00 2.15	57.1	31.9	30.0
2	1.158	0.523	0.447	0.240	0.356	0.305	0.86 13.97	0.00 1.98	0.00 1.53	60.3	23.0	21.4
3	1.309	0.982	0.750	0.136	0.347	0.261	0.88 2.03	0.00 2.03	0.00 1.54	62.9	35.4	29.0
4	1.189	0.614	0.535	1.890	0.384	0.341	0.82 103.60	0.00 1.96	0.00 1.44	61.5	22.5	20.6
5	1.300	1.000	0.767	0.122	0.297	0.219	0.88 1.50	0.00 2.10	0.00 1.58	61.6	36.2	30.8
6	0.569	0.490	0.904	0.270	0.170	0.365	0.27 7.92	0.04 1.03	0.00 2.86	14.1	7.5	22.7
7	0.972	0.925	0.956	0.057	0.117	0.142	0.67 1.05	0.52 1.24	0.51 1.50	55.3	37.7	35.8
8	0.791	0.733	0.940	0.101	0.136	0.207	0.38 0.99	0.20 1.09	0.42 1.79	18.8	17.1	30.6
9	1.018	0.976	0.960	0.030	0.090	0.100	0.82 1.06	0.66 1.29	0.64 1.48	62.0	41.9	36.8
10	1.256	0.841	0.671	0.217	0.388	0.306	0.87 4.41	0.00 2.09	0.00 1.56	66.0	29.3	25.2
11	0.916	0.865	0.951	0.076	0.129	0.166	0.64 1.04	0.45 1.34	0.45 1.63	36.6	30.2	34.4
12	0.974	0.851	0.885	0.120	0.191	0.225	0.48 1.14	0.25 1.35	0.31 2.23	53.8	32.9	32.4
13	0.906	0.752	0.841	0.150	0.235	0.284	0.50 1.13	0.01 1.42	0.01 2.08	39.2	22.9	28.1
14	1.018	0.912	0.903	0.101	0.170	0.190	0.39 1.13	0.32 1.40	0.37 2.03	58.6	34.7	32.0

CLS = Complete least squares (before censoring); ILS = Iterative least squares; ML = Maximum likelihood

TABLE 7A  
DETAILED COMPARISON OF RESULTS FROM 5000 SIMULATION RUNS FOR ESTIMATED REGRESSION INTERCEPT  
AND MEAN AT LOW STRESS ( $\beta_0 = 1$ ) AT TEST CONDITION 1: ITERATIVE LEAST SQUARES VERSUS MAXIMUM LIKELIHOOD

Iterative Least Squares Estimate of $\beta_0$	Maximum Likelihood Estimate of $\beta_0$												Total	
	Upper Limit of Frequency Class													
	0.10	0.30	0.50	0.70	0.90	1.10	1.30	1.50	1.70	1.90	$\infty$			
0.10	24	0	0	0	0	0	0	0	0	0	0	0	24	
0.30	0	95	0	0	0	0	0	0	0	0	0	0	95	
0.50	0	8	306	3	0	0	0	0	0	0	0	0	317	
0.70	0	0	45	642	7	0	0	0	0	0	0	0	694	
0.90	0	0	0	70	964	35	1	0	0	0	0	0	1070	
1.10	0	0	0	0	79	859	88	4	3	0	0	0	1033	
1.30	0	0	0	0	0	14	589	133	10	2	1	1	749	
1.50	0	0	0	0	0	0	5	261	148	21	2	2	437	
1.70	0	0	0	0	0	0	0	0	133	87	42	42	262	
1.90	0	0	0	0	0	0	0	0	0	38	85	85	123	
$\infty$	0	0	0	0	0	0	0	0	0	0	196	196	196	
Total	24	103	351	715	1050	908	683	398	294	148	326	326	5000	

TABLE 7C  
DETAILED COMPARISON OF RESULTS FROM 5000 SIMULATION RUNS FOR ESTIMATED REGRESSION INTERCEPT  
AND MEAN AT LOW STRESS ( $\beta_0 = 1$ ) AT TEST CONDITION 1: MAXIMUM LIKELIHOOD VERSUS COMPLETE LEAST SQUARES\*

Maximum Likelihood Estimate of $\beta_0$	Complete Least Squares Estimate of $\beta_0$													Total	
	Upper Limit of Frequency Class		Upper Limit of Frequency Class												
	0.10	0.30	0.50	0.70	0.90	1.10	1.30	1.50	1.70	1.90	$\infty$				
0.10	16	8	0	0	0	0	0	0	0	0	0	0	0	24	
0.30	5	59	28	11	0	0	0	0	0	0	0	0	0	103	
0.50	0	16	162	122	41	7	2	1	0	0	0	0	0	351	
0.70	0	0	52	275	271	92	17	8	0	0	0	0	0	715	
0.90	0	1	14	121	397	311	152	40	13	1	0	0	0	1050	
1.10	0	0	2	36	170	305	262	91	37	5	0	0	0	908	
1.30	0	0	1	10	71	196	226	131	37	11	0	0	0	683	
1.50	0	0	1	3	30	95	121	90	39	16	3	0	0	398	
1.70	0	0	0	3	17	49	96	81	34	10	4	0	0	294	
1.90	0	0	0	0	4	18	37	42	39	6	2	0	0	148	
$\infty$	0	0	0	1	5	36	76	102	61	31	14	0	0	326	
Total	21	84	260	582	1006	1109	989	586	260	80	23	0	0	5000	

\* Before censoring

TABLE 7B  
DETAILED COMPARISON OF RESULTS FROM 5000 SIMULATION RUNS FOR ESTIMATED REGRESSION INTERCEPT  
AND MEAN AT LOW STRESS ( $\beta_0 = 1$ ) AT TEST CONDITION 1: ITERATIVE LEAST SQUARES VERSUS  
COMPLETE LEAST SQUARES\*

Iterative Least Squares Estimate of $\beta_0$	Upper Limit of Frequency Class	Complete Least Squares Estimate of $\beta_0$											Total
		Upper Limit of Frequency Class											
		0.10	0.30	0.50	0.70	0.90	1.10	1.30	1.50	1.70	1.90	$\infty$	
0.10	16	8	0	0	0	0	0	0	0	0	0	0	24
0.30	5	56	26	8	0	0	0	0	0	0	0	0	95
0.50	0	19	150	112	30	5	1	0	0	0	0	0	317
0.70	0	0	63	278	248	81	17	7	0	0	0	0	694
0.90	0	1	16	128	417	296	153	42	16	1	0	0	1070
1.10	0	0	3	42	196	363	283	102	39	5	0	0	1033
1.30	0	0	2	10	78	203	249	148	46	13	0	0	749
1.50	0	0	0	2	25	90	141	112	42	20	5	0	437
1.70	0	0	0	0	9	43	72	75	46	11	4	0	262
1.90	0	0	0	0	3	19	32	38	28	2	1	0	123
$\infty$	0	0	0	0	0	9	41	62	43	28	13	0	196
Total	21	84	260	582	1006	1109	989	586	260	80	23	0	5000

\* Before censoring

TABLE 8  
OVERALL COMPARISON OF THE PERFORMANCE OF THE ITERATIVE LEAST SQUARES  
AND MAXIMUM LIKELIHOOD ESTIMATION METHODS AT THE 14 TEST CONDITIONS  
IN THE SIMULATION ANALYSIS

Number of Conditions Out of 14:	Intercept and Mean at Low Stress ( $\beta_0$ )	Slope ( $\beta_1$ )	Mean at High Stress ( $\beta_0 + \beta_1$ )	Standardized Residual Error Standard Deviation ( $\sigma'$ )
1) For which average ILS estimate is closer to true parameter value than is ML estimate	7	6*	4*	7
2) For which ILS estimate has lower root mean square error than ML estimate	12	13	9*	8
3) For which ILS estimate has more simulation runs closer to true parameter value than ML estimate has	10	10	**	10
Average percentage of runs (over 14 conditions) for which ILS estimate is closer to true parameter value than ML estimate	63.8%	56.2%	**	50.6%

\* Out of 13 conditions.  
\*\* Not calculated.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Simple Method For Regression Analysis With Censored Data		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6/1/78-4/15/79
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Josef Schmee, Gerald J. Hahn		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0438
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute of Administration & Management, Union College and University, Schenectady, NY 12308		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of Navy Arlington, VA 22217		12. REPORT DATE April 15, 1979
		13. NUMBER OF PAGES 52
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES To be presented in TECHNOMETRICS Session at ASA Annual Meeting, Washington, DC, 1979; to be published in TECHNOMETRICS, November 1979.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Censored data; regression analysis; maximum likelihood; accelerated testing; life testing; least squares.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Problems requiring regression analysis of censored data arise frequently in practice. For example, in accelerated testing one wishes to relate stress and average time to failure from data including unfailed units, i.e., censored observations. Maximum likelihood is one method for obtaining the desired estimates; in this paper, we propose an alternative approach. An initial least squares fit is obtained treating the censored (over)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

values as failures. Then, based upon this initial fit, the expected failure time for each censored observation is estimated. These estimates are then used, instead of the censoring times, to obtain a revised least squares fit and new expected failure times are estimated for the censored values. These are then used in a further least squares fit. The procedure is iterated until convergence is achieved. This method is simpler to implement and explain to non-statisticians than maximum likelihood and appears to have good statistical and convergence properties.

The method is illustrated by an example, and some simulation results are described. Variations and areas for further study also are discussed.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)